Time Series Analysis of Melbourne’s Temperature

• Executive summary

The debate on global warming has drawn the public’s attention. In this report, time series analysis is performed on the average temperature in Melbourne over the past 5 decades with an aim to find out the temperature pattern and make predictions for the next 10 years using fitted model. The result shows a stable temperature pattern from 1970 to 2000 in the data but an obvious increasing trend from 2000. Therefore, if no actions are taken to address this issue, the increasing trend tends to continue for the following 10 years.

• Discussions of the issues on hand and any exploratory data features

- ***Dataset Description and Issues on hands***

The dataset describes Melbourne’s average air temperature measured from July 1970 to February 2020, which contains information about year, month, minimum and maximum temperatures.

Monthly average temperatures are then calculated using the average of minimum and maximum at each time point. The aim of this repost is to use the monthly average temperature to explore the weather pattern over the past 50 years and make a prediction on the temperature pattern in the next 10 years.

- ***Exploration data features***

There are no missing values in the dataset. As shown in the time series plot (Appendix 1), there is a frequently fluctuated pattern in the temperature during 1970 to 2020 and the average temperatures range between 7.8 degree Celsius to 23.55 degree Celsius. Since the temperature is recorded on a monthly basis, a centred moving average of order 12 is then performed.

Author: Sally Chen

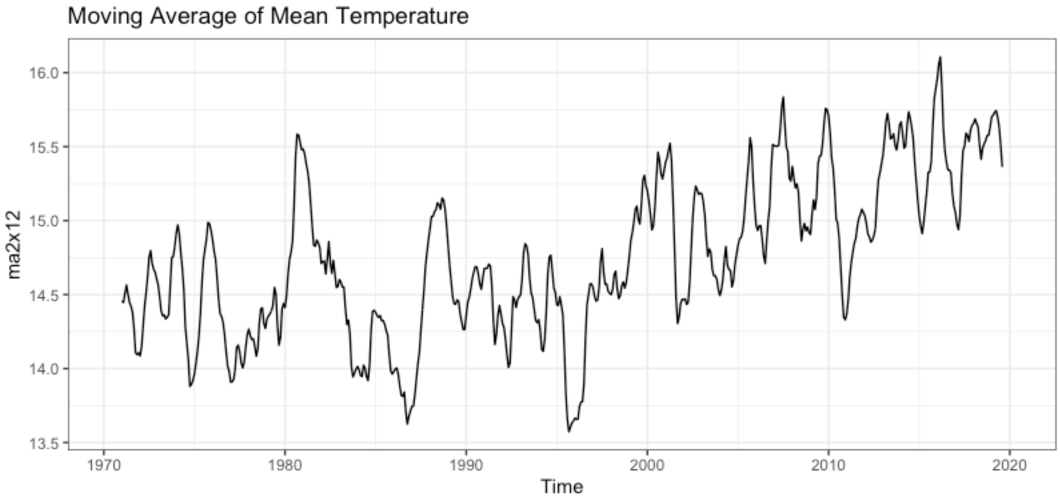


Figure 1. Moving Average of Mean Temperature

In the plot (Figure 1) above, the moving average of temperature fluctuated around a similar value before year 2000 with a peak point at year 1981 and bottom points at 1988 and 1996. After 2020, there is an overall increasing trend with a short drop around 2010. The similar trend pattern is also captured by the combination plot of backward moving average at order 12, 24, 48 (Appendix2) which can avoid the loss of the most recent data information, therefore can be used as a good supplement tool.

• Technical analysis

STL decomposition is first performed on the dataset with manually selected t.window (smoothness of the trend-cycle) and s.window (rate of change of seasonal component). Acf plot and Box-Ljung test are used to test the white noise property of the decomposition. After that, seasonally adjusted data is extracted for further regression analysis and forecast evaluation.

In order to test predictive and forecast accuracy of the potential models, the seasonally adjusted dataset is then split into training set and test set. Piecewise linear regression with different turning points are applied to model the training set.

To compare models' prediction performance, cross validation tests are performed and statistics values of AIC, BIC and Adj R^2 are used. To compare models' forecast accuracy, accuracy test is performed, and difference error statistics including RMSE, MAE, MAPE and MASE are used. Besides, Diebold& Mariano test is followed to compare the significant difference of forecast error between two models

• Discussion of key results

○ *STL Decomposition of the Data*

STL decomposition with t.window = 12 and s.window = 6 is performed and the trend, seasonal and remainder components are plotted separately along with the original time series plot (Figure2). The trend component is smoother compared to the original data, the seasonal component is changing slowly over time, and there is a random pattern in the remainder.

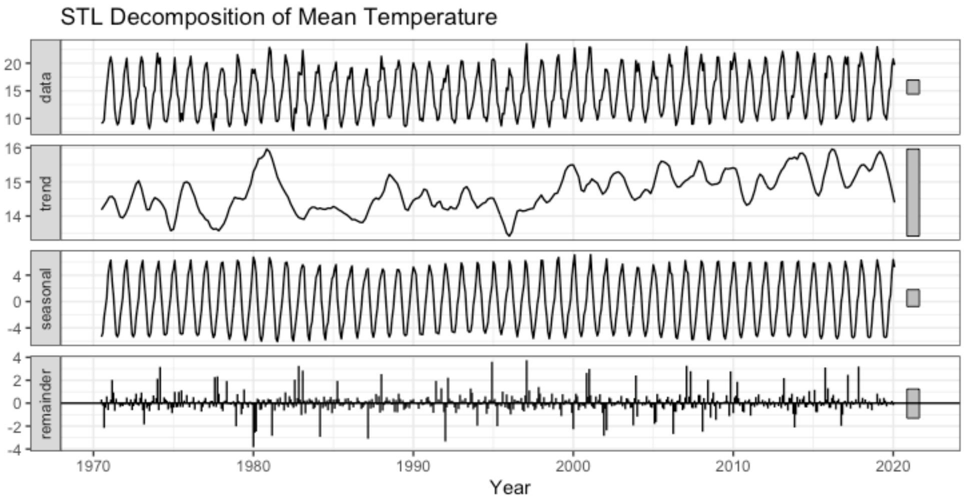


Figure 2. STL Decomposition of Mean Temperature

▪ **Seasonal Component Analysis**

To have a closer analysis of the seasonal component, a Monthplot is generated (Figure 3). The Monthplot summarises the different fluctuation patterns and different values of the seasonal component through the months, during summer (January – March) the values of seasonal component are positive while during winter (June – August) the values of seasonal component are negative.

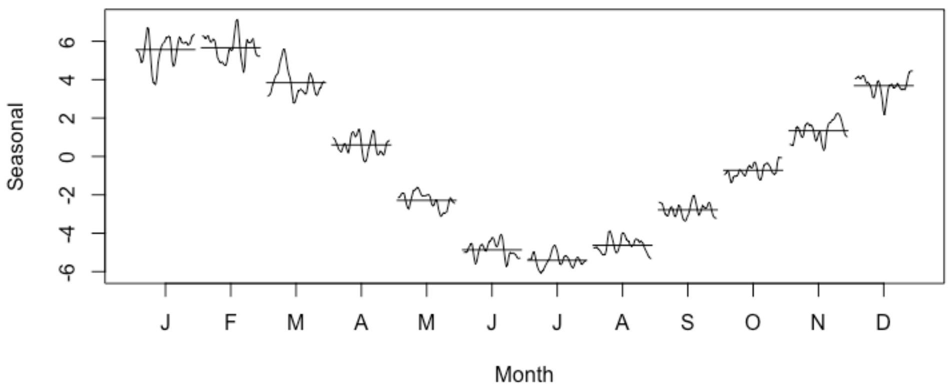


Figure3. Monthplot of Seasonal Component

- **Remainder Component Diagnosis**

Acf plot of the remainder component (Figure 4) is generated to estimate the autocorrelation of the time series. Most of the acf values are within the dot line, indicating the white noise property and therefore a good decomposition. Box-Ljung test also indicates the result that the autocorrelations are jointly equal to 0 (Appendix 3).

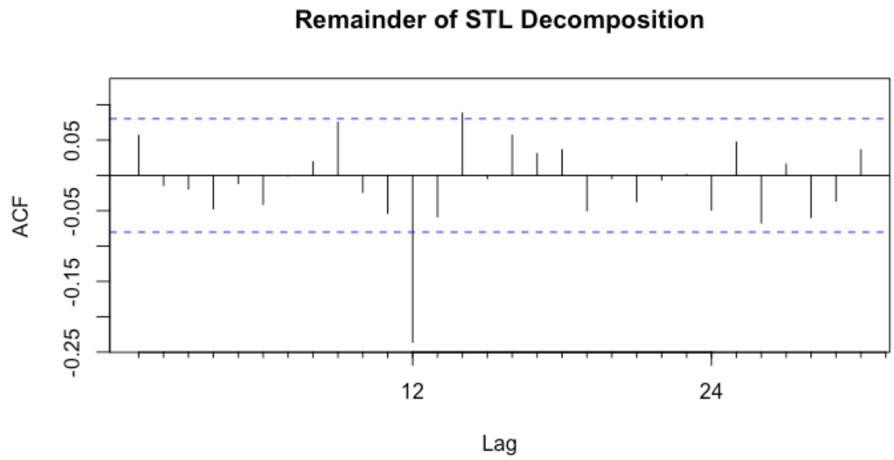


Figure 4. Remainder of STL Decomposition

- ***Regression Analysis***

After STL decomposition, seasonally adjusted data is extracted and is then separated into training set (start – February 2010) (Appendix 4) and test set (March 2020 – end) (Appendix 5). Piecewise linear regressions with different turning points are used to fit regression models.

- **Piecewise Linear Regression with Four Turning Points (PLR4)**

As we can see in the time series plot of training set (Appendix 5), there are 4 turning points, therefore a piecewise linear regression is fitted:

$A_t = 14.2480 + 0.0035 trend_i - 0.0122 trend1_i^+ + 0.0150 trend2_i^+ - 0.0122 trend3_i^+ + 0.0123 trend_i^4 + \epsilon_i$,

with $trend1_i^+ = \max(trend_i - 127, 0)$, $trend2_i^+ = \max(trend_i - 190, 0)$, $trend3_i^+ = \max(trend_i - 263,0)$, $trend4_i^+ = \max(trend_i - 308,0)$.

Refer to Appendix 6 for the plot of fitted value and seasonally adjusted data.

- **Piecewise Linear Regression with Two Turning Points (PLR2)**

Since the first two turning points are more obvious than the rest (Appendix 4), to reduce model complexity, the fitted model can be changed to

$A_t = 14.2474 + 0.0035trend_i - 0.0123trend1_i^+ + 0.0127trend2_i^+ + \epsilon_i$, with $trend1_i^+ = \max(trend_i - 127, 0)$, $trend2_i^+ = \max(trend_i - 190, 0)$.

Refer to Appendix 7 for the plot of fitted value and seasonality adjusted data.

- **Model Comparison**

  PLR4 shows a better fit for the data in the regression plots (Appendix 6&7). Besides, Cross Validation diagnoses are performed for the two models and the results are shown below:

PLR4	CV	AIC	BIC	AdjR2
	0.94355662	-25.83388344	3.32404154	0.08365554
PLR2	CV	AIC	BIC	AdjR2
	0.94522594	-24.96822504	-4.14113576	0.07816336

Table 1. Model Fit Measures

PLR2 has a low BIC while PLR4 has a lower cross validation error together with lower AIC and higher adjusted R^2. Thus, PLR4 is a better fitted model for predictive accuracy.

- ***Forecast Evaluation***

After model fit on the test set, 120 steps forecasts are performed on the training set using two models, the seasonality components are added back to the fitted value and the forecast accuracies are tested using the test set (with seasonality components added back).

The forecast plots of both regressions show an increasing trend (Appendix 8). The result of accuracy test shows that PLR2 has an overall smaller forecast error.

PLR4	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	3.948894	3.4397866	-7.953892	25.88923	2.9894381	0.8128762
Test set	1.018083	0.8437648	-4.101302	5.90031	0.7332962	0.3064732
PLR2	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	3.9506512	3.4431170	-7.959386	25.911792	2.9923324	0.8128910
Test set	0.8915683	0.6825558	-1.722182	4.642317	0.5931932	0.3088086

Table 2. Forecast Accuracy Measures of Two Regression Models

One-sided Diebold-Mariano Test is also performed to test the error difference of two models and the result shows that PLR2 has a smaller error than PLR4 with p-value close to 0, which supports the superior performance of PLR2 in forecasting.

• Conclusion and limitations

Based on the analysis of the average temperature pattern of Melbourne over the past 40 years, it is concluded that although there were some drastic change points and fluctuation in the mean temperature from year 1970 to 2000, there is no obvious increasing trend. However, during the past 20 years, the increasing trend has become obvious, which aligns with the effect of global warming.

Since PLR4 performs better with in-sample prediction and PLR2 performs better for out-of-sample forecasting, the latter is used to forecast Melbourne's average temperature for the next 10 years. The plot (Figure 4 indicates an increasing trend of the mean temperature with an average increase of 1-2 degree Celsius.

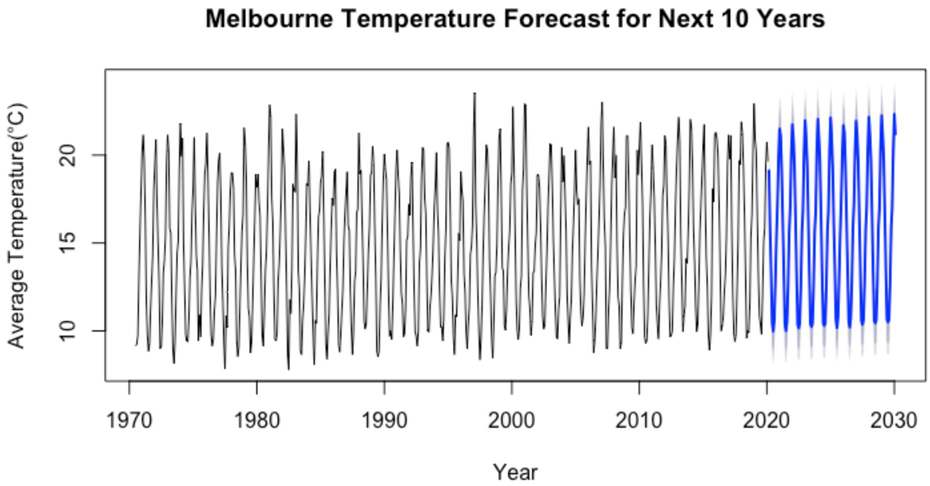


Figure 4. Melbourne Temperature Forecast for Next 20 Years

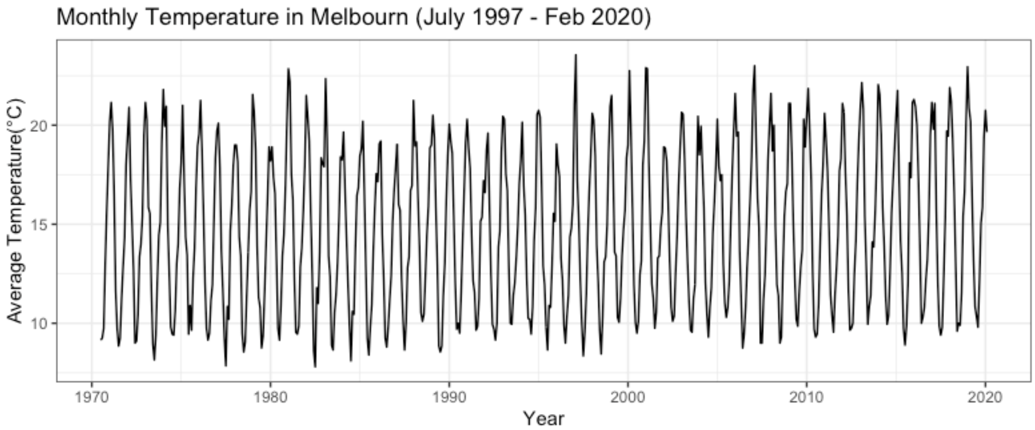
However, the limitations of the model also need to be considered.

In terms of model fit accuracy, the nature of the dataset leads to large values of the remainder component during decomposition, and it can result in the poor capture of the model trend when fitting regression. Besides, the autocorrelation tests (Appendix 9) shows there are still some pattern in the remainder that are not captured by the models, therefore more techniques should be applied to find a better model fit.

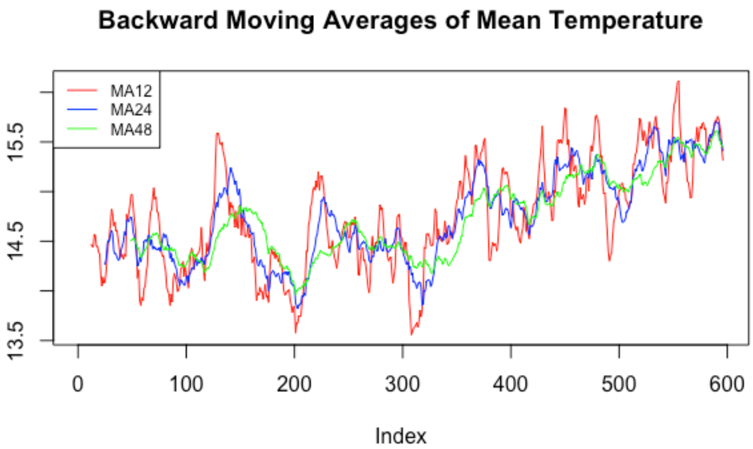
In terms of forecast accuracy, naïve forecast of the seasonal component is used here, but due to climate change and other natural effects, the seasonal effect tends to be changed over time.

Since the increasing trend of temperature is foreseeable, it is suggested that individuals, government and organisations should take actions to reduce the effect of climate change on future generations.

• **Appendix**



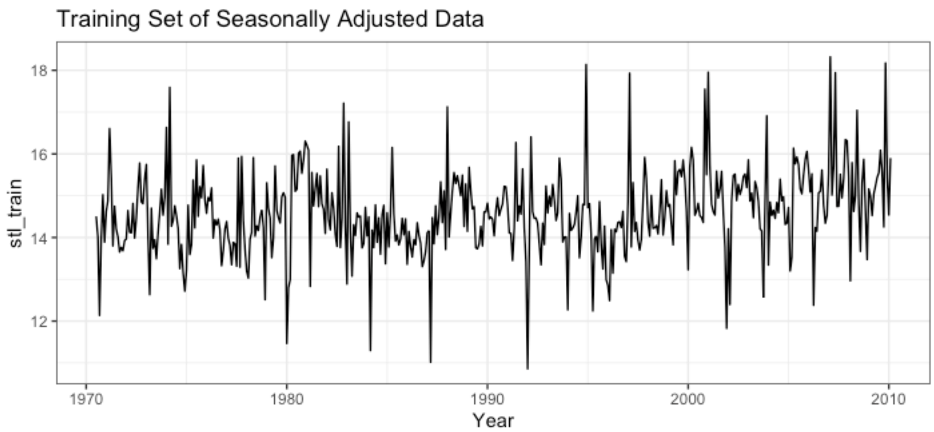
Appendix 1. Monthly Temperature in Melbourne (July 1997 – Feb 2020)



Appendix 2. Backward Moving Average of Mean Temperature

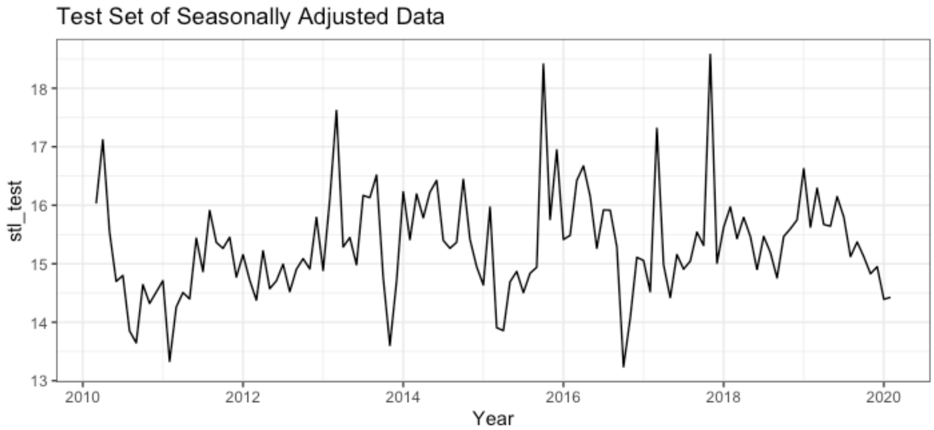
Box-Ljung test		
X-squared = 1.9062	df = 1	p-value = 0.1674

Appendix 3. Box-Ljung test result

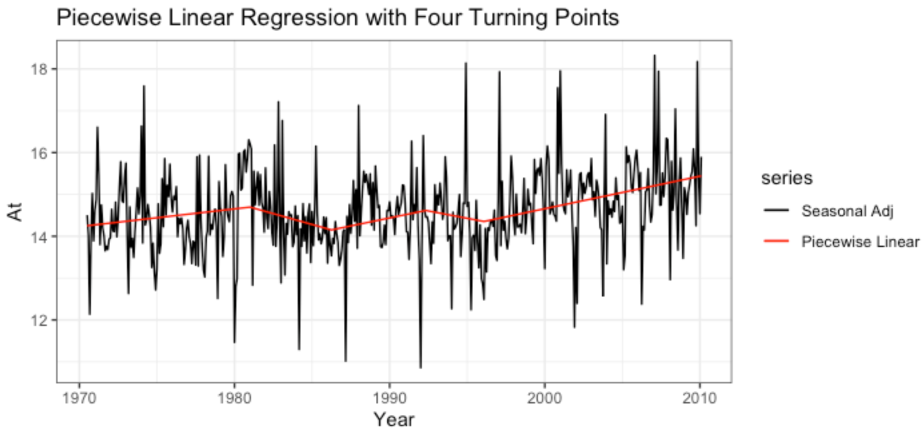


Appendix 4. Training Set of Seasonally Adjusted Data

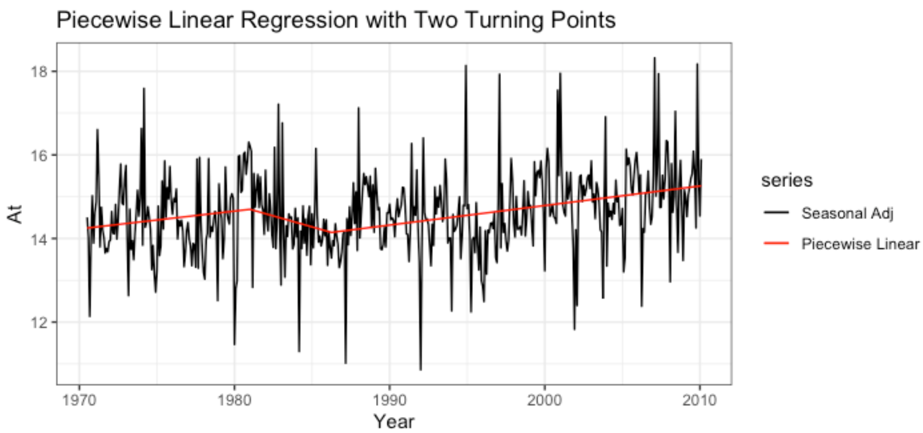
Author: Sally Chen



Appendix 5. Test Set of Seasonally Adjusted Data

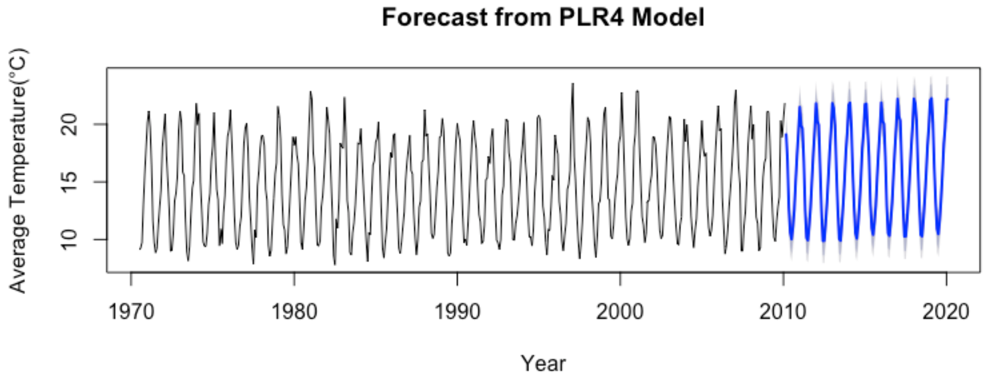


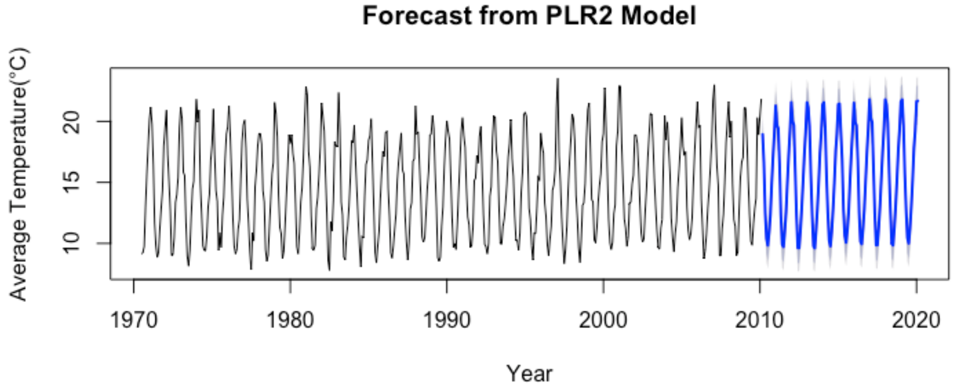
Appendix 6. Fitted Value of PLR4 and Seasonally Adjusted Data



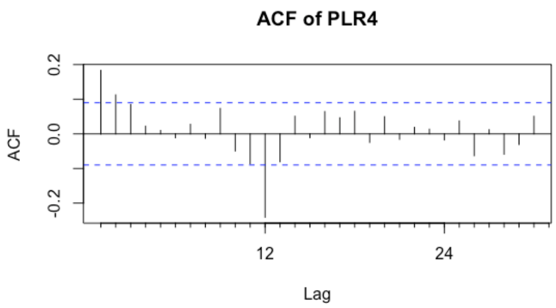
Appendix 7. Fitted Value of PLR2 and Seasonally Adjusted Data

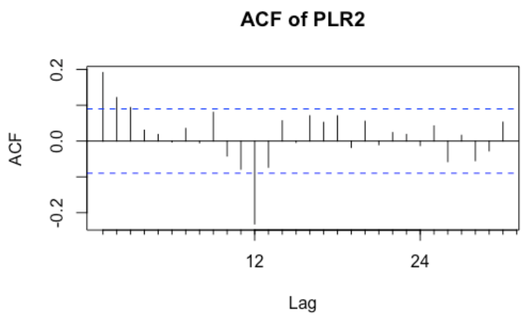
Author: Sally Chen





Appendix 8. Forecast Plots of Two Regression Models





Appendix 9. Acf plots of Two Regression Models

Author: Sally Chen